

Dados geográficos e inteligência artificial na predição de locais favoráveis para a instalação de usinas de biogás da agroindústria canavieira

The use of geographic data and artificial intelligence to predict suitable locations for biogas plant installation in the sugarcane agroindustry

Données géographiques et intelligence artificielle sont utilisées pour anticiper les emplacements propices à l'installation d'usines de biogaz dans l'agroindustrie de la canne à sucre



Marlísia D'Abadia de Pina

Instituto Federal de Educação, Ciência e Tecnologia de Goiás –
Goiânia-Goiás-Brasil

pina.marlisia@academico.ifg.edu.br



Édipo Henrique Cremon

Instituto Federal de Educação, Ciência e Tecnologia de Goiás –
Goiânia-Goiás-Brasil

edipo.cremon@ifg.edu.br

Resumo A Agenda 2030 da ONU visa promover o aumento das energias renováveis em todo o mundo, e a biomassa agrícola, em particular, a partir da cana-de-açúcar, é uma solução relevante para essa transição energética no Brasil. A localização geográfica desempenha um papel crucial na determinação do local ideal para a instalação de usinas de biogás, e a combinação de dados geográficos e algoritmos de Inteligência Artificial, conhecida como GeoIA, oferece

uma abordagem promissora para prever esses locais ideais. Nesse sentido, este estudo teve por objetivo prever locais favoráveis para a instalação de usinas de biogás proveniente da cana-de-açúcar da agroindústria, utilizando dados geográficos associados a aspectos físicos, bióticos e antrópicos, além de seis tipos de algoritmos de classificação (CART, C4.5, C5.0, Random Forest, XGBoost e GBM) para comparar seus desempenhos. O treinamento foi feito para o estado de São Paulo, devido ao número maior de usinas contidas na unidade federativa, e o modelo de melhor desempenho foi aplicado para o estado de Goiás. O algoritmo Random Forest obteve o melhor desempenho e permitiu identificar locais favoráveis para a instalação de usinas de biogás de cana-de-açúcar em Goiás. Essa abordagem pode facilitar a tomada de decisões ao identificar regiões propícias para a produção de biogás a partir de cana-de-açúcar, otimizando o uso de biomassa, reduzindo o impacto ambiental e os custos de instalação. O uso de GeolA contribui para a expansão das energias renováveis e a mitigação das mudanças climáticas, promovendo a transição energética global.

Palavras-Chave: Aprendizado de máquina; Biomassa; Energia renovável; Sistema de informações geográficas.

Abstract Aligned with the UN 2030 Agenda's imperative to facilitate the widespread adoption of renewable energies, this study underscores the pertinence of agricultural biomass, notably derived from sugarcane, as a substantive solution to Brazil's ongoing energy transition. The determination of optimal sites for the deployment of biogas plants is inherently contingent upon geographic considerations. This research advocates for the integration of geographic data with Artificial Intelligence algorithms, colloquially termed GeolA, as a robust and prospective methodology for the precise anticipation of these optimal locations. In consideration of the foregoing, this study endeavors to forecast optimal sites for the implementation of sugarcane biogas plants within the agro-industry. Leveraging geographical data encompassing physical, biotic, and anthropic facets, alongside the employment of six distinct classification algorithms (CART, C4.5, C5.0, Random Forest, XGBoost, and GBM), performance comparison becomes paramount. The training phase specifically targeted the state of São Paulo, owing to its heightened concentration of plants, with the most efficacious model subsequently applied to the state of Goiás. The preeminent performance achieved by the Random Forest algorithm underscores its efficacy in delineating advantageous sites for the deployment of sugarcane biogas plants in Goiás. This methodological approach holds promise in streamlining decision-making processes, delineating regions conducive to biogas production from sugarcane, thereby optimizing biomass utilization and concurrently mitigating environmental impact and installation expenditures. The incorporation of GeolA not only fosters the proliferation of renewable

energies but also contributes substantively to climate change mitigation, thereby catalyzing the broader global energy transition.

Keywords: Machine learning; Biomass; Renewable energy; Geographic information system.

Resume L'Agenda 2030 des Nations Unies vise à promouvoir l'essor des énergies renouvelables à l'échelle mondiale. La biomasse agricole, notamment celle issue de la canne à sucre, apparaît comme une solution pertinente pour la transition énergétique au Brésil. La localisation géographique revêt un rôle crucial dans la détermination des sites optimaux pour l'implantation de centrales de biogaz. La combinaison de données géographiques et d'algorithmes d'Intelligence Artificielle, regroupés sous le terme de GeolA, offre une approche prometteuse pour anticiper ces emplacements privilégiés. Dans cette perspective, l'objectif de cette étude était de prédire les sites favorables à l'installation de centrales de biogaz issues de la canne à sucre dans l'agro-industrie. Ceci a été réalisé en utilisant des données géographiques associées à des aspects physiques, biotiques et anthropiques, ainsi que six types d'algorithmes de classification (CART, C4.5, C5.0, Random Forest, XGBoost et GBM) afin de comparer leurs performances. L'entraînement a été effectué dans l'État de São Paulo, en raison du nombre prépondérant d'installations dans cette unité fédérative, et le modèle présentant les meilleures performances a été appliqué à l'État de Goiás. L'algorithme Random Forest a démontré les meilleures performances, permettant ainsi d'identifier les sites favorables à l'implantation de centrales de biogaz issues de la canne à sucre à Goiás. Cette approche facilite la prise de décision en identifiant les régions propices à la production de biogaz à partir de la canne à sucre, optimisant ainsi l'utilisation de la biomasse tout en réduisant l'impact environnemental et les coûts d'installation. L'utilisation de GeolA contribue à l'expansion des énergies renouvelables et à l'atténuation du changement climatique, favorisant ainsi la transition énergétique à l'échelle mondiale.

Mots clés: Apprentissage automatique; Biomasse; Énergie renouvelable; Système d'information géographique

Introdução

A sustentabilidade tornou-se uma questão cada vez mais proeminente no contexto mundial, pois implica a busca pelo equilíbrio entre a exploração dos recursos naturais e a preservação do meio ambiente. Em 2015, a Organização das Nações Unidas (ONU) lançou a Agenda 2030, composta por 17 Objetivos de Desenvolvimento Sustentável (ODS). O tópico da energia acessível e limpa é abordado no ODS número 7, que tem como objetivo aumentar substancialmente a participação das energias renováveis na matriz energética global, duplicar as taxas de melhoria da eficiência energética e promover o acesso à pesquisa, tecnologia e investimento em energia limpa, além de expandir e modernizar os serviços energéticos em países em desenvolvimento (ONU, 2015).

Nesse contexto, o setor energético está passando por uma revolução, com a produção de biocombustíveis, por meio de tecnologias limpas, que substituem o uso de combustíveis fósseis e reduzem significativamente as emissões de Gases de Efeito Estufa (GEE) (Romero et al., 2023). Além disso, a exploração de recursos energéticos não renováveis tem uma duração limitada. Conforme evidenciado por estudos, as reservas de petróleo, carvão e gás estarão esgotadas em torno de 35, 107 e 37 anos, respectivamente, levando em consideração a taxa atual de consumo e as projeções de demanda energética (Ioannou et al., 2018). Portanto, torna-se cada vez mais importante a utilização de fontes renováveis, que não só mitigam os efeitos climáticos, mas também derivam de recursos naturais que se renovam em um período de tempo compatível com a vida humana e estão disponíveis para uso sem se esgotarem.

A participação da energia renovável no mercado está em crescimento, impulsionada por bioprodutos como bioetanol, biodiesel e

biogás, derivados da biomassa agrícola (Romero et al., 2023). A agroindústria canaveira é uma fonte considerável de resíduos que podem ser convertidos em biogás. O Valor Bruto da Produção (VBP) da cana-de-açúcar atingiu a marca de 80 bilhões de reais em 2022 (CNA, 2023), representando quase 1% do PIB brasileiro, estimado em 9,9 trilhões de reais (IBGE, 2023). Com uma produção tão grande, essa indústria possui um imenso potencial de resíduos orgânicos, o que a torna uma lucrativa fonte de matéria-prima.

O uso do biogás como fonte energética é visto como uma opção viável a fontes convencionais, pois é uma matriz energética capaz de promover o desenvolvimento ecológico, econômico e social sustentável, podendo ser produzido a partir de diversas fontes de resíduos renováveis.

Além de contribuir para o combate às mudanças climáticas, o biogás é útil para a distribuição descentralizada de energia e uma tecnologia adaptável a diferentes escalas de produção e ambientes geográficos, bem como pode ser utilizado para diversas finalidades energéticas (Jende et al., 2016). Sua utilização desempenha um papel importante na economia circular, não só pelo fornecimento de energia renovável, mas também por ser identificado como um processo regenerativo e restaurador, que elimina a geração de resíduos ao reutilizar e reciclar matérias-primas e seus próprios resíduos (Romero et al., 2023).

A determinação de locais ideais para a instalação de usinas de biogás ainda é um desafio complexo, devido à variedade de fatores a serem considerados, como a proximidade aos consumidores, a disponibilidade de matéria-prima, as condições de infraestrutura e os aspectos regulatórios (Ioannou et al., 2018). Nesse contexto, a associação entre técnicas de inteligência artificial (IA), como o aprendizado de máquina (machine learning), e dados geográficos pode otimizar o planejamento do uso do biogás, contribuindo para o desenvolvimento sustentável.

A partir da combinação entre IA e dados geográficos, surge a GeolA, um subcampo da ciência de dados espaciais que oferece suporte para a extração de informações geográficas de maneira mais inteligente (Janowicz et al., 2020). Sendo assim, a GeolA pode otimizar o uso do biogás, auxiliando na indicação de locais favoráveis para a instalação de usinas de biogás, uma vez que permite uma análise mais completa e precisa. Isso ocorre pela combinação de informações ambientais, sociais e econômicas com dados geográficos, o que gera modelos precisos para promover o desenvolvimento sustentável (Romero et al., 2023).

A utilização de algoritmos no contexto da GeolA pode ser considerada uma ferramenta de grande valor na predição de locais promissores para a instalação de usinas de biogás. Com a capacidade de analisar múltiplas variáveis de maneira simultânea, bem como de aprender continuamente com base em dados existentes, tais algoritmos podem gerar modelos mais precisos e confiáveis para orientar as decisões dos gestores.

Esta pesquisa busca, assim, alternativas de bioenergia e a minimização de impactos ambientais. O biogás se mostra uma opção viável para a geração de energia limpa e contribuição para o desenvolvimento sustentável. Dessa forma, este estudo tem por objetivo explorar o potencial do uso de algoritmos de aprendizado de máquina, combinado com dados geográficos para a predição de locais favoráveis à instalação de usinas de biogás na agroindústria canavieira.

Metodologia

Neste estudo, oito variáveis foram definidas para predição de locais propícios à instalação de usinas de biogás da agroindústria canavieira (Tabela 1). Essas variáveis foram escolhidas pautadas em revisão bibliográfica.

Tabela 1: Relação das variáveis utilizadas para predição de locais favoráveis à instalação de usina de biogás derivado de resíduos de cana-de-açúcar e fonte bibliográfica.

| Variáveis/Autores | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| Distância da área produtiva (agrícola e silvicultura) | | | | X | | X | | | X | |
| Distância da área urbana | | X | X | X | X | X | X | X | X | X |
| Distância de áreas com cana-de-açúcar | X | | | X | X | X | X | X | X | X |
| Distância de rodovias e ferrovias | X | X | X | X | X | X | X | X | X | X |
| Distância de vegetação nativa | X | X | X | X | X | X | X | X | X | X |
| Distância de hidrografia | | X | X | X | | X | X | X | X | X |
| Distância de trecho de energia | X | X | X | X | X | X | | X | X | X |
| Declividade | | | X | X | | X | | X | X | X |

Legenda: A) Dagnall, Hill e Pegg (2000); B) Sliz-Szkliniarz e Vogt (2012); C) Sultana e Kumar (2012); D) Silva, Alçada-Almeida e Dias (2014); E) Franco et al. (2015); F) Ioannou et al. (2018); G) Laasasenaho et al. (2019); H) Lozano-García et al. (2020); I) Yalcinkaya (2020); J) Zhao et al. (2022).

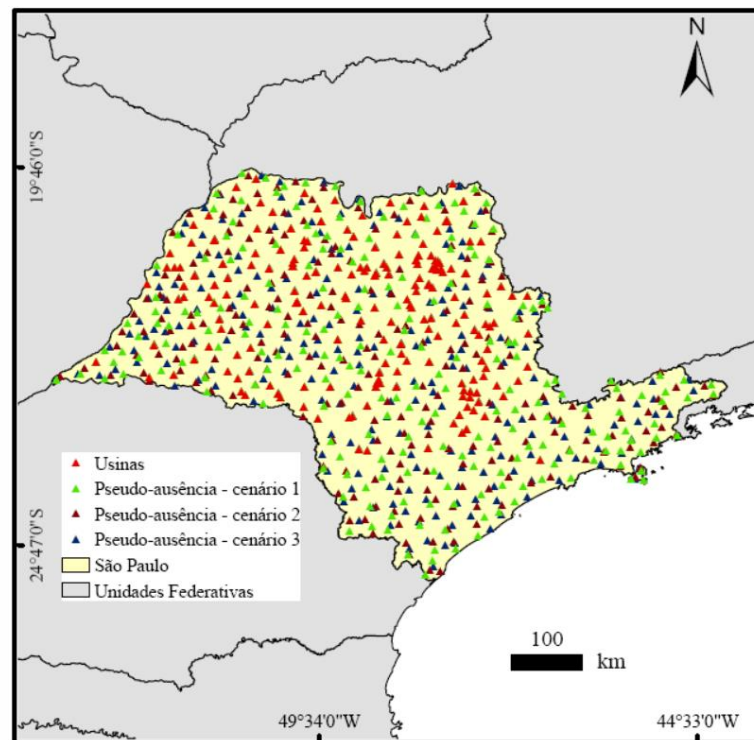
Fonte: Autoria própria.

Para o treinamento de um modelo de inteligência artificial, foram considerados os dados referentes ao estado de São Paulo, situado na região sudeste do Brasil, devido à sua concentração significativamente maior de usinas de energia derivada da cana-de-açúcar (um total de 214 unidades) em relação às demais unidades da federação no Brasil.

Os dados de localização das usinas foram adicionados como camada vetorial de pontos por meio do software livre QGIS (versão 3.28). Esse procedimento teve por intuito gerar um arquivo que pudesse, posteriormente, agregar para cada um desses pontos os valores de pixel das camadas matriciais usadas na predição (Tabela 1). Foram gerados, de forma aleatória, 215 pontos de pseudo-ausência (onde não há usinas). A definição de pontos de pseudo-ausência é frequente em modelagens que possuem informações sobre a presença espacial de um fenômeno ou alvo a ser predito, mas desconhecem a sua ausência (Zurell et al., 2020).

Para uma análise mais aprofundada da distribuição dos pontos considerados como pseudo-ausência, foram criadas três camadas vetoriais distintas, denominadas cenário 1, cenário 2 e cenário 3 (Figura 1).

Figura 1: Usinas de energia da cana e pontos de pseudo-ausência, considerados nos três cenários.



Fonte: Autoria própria.

Em relação à base de dados geográficos, para o estado de São Paulo, as distâncias de área produtiva (agrícola e silvicultura), cana-de-açúcar, área urbana e vegetação nativa foram considerados os dados do MapBiomass coleção 8, relativos ao ano de 2020 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil (Souza et al., 2020). Para a distância em relação à hidrografia, trecho de energia e rodovias e ferrovias foram considerados os dados do Departamento de Águas e Energia Elétrica (DAEE) (2008) e do Open Street Map (OSM, 2023). A partir desses dados foram geradas as distâncias euclidianas em formato matricial (raster) em ambiente de Sistema de Informação Geográfica (SIG), com tamanho de pixel de 30m. Por

sua vez, a declividade, com esse mesmo tamanho de pixel, foi obtida a partir do Modelo Digital de Elevação Copernicus. Todos os dados foram projetados em coordenadas métricas na projeção cônica equivalente de Albers.

Para este estudo, seis tipos de algoritmos foram utilizados, sendo eles: CART (Classification and Regression Trees), C4.5, C5.0, Random Forest, XGBoost e GBM (Gradient Boosting Machine). O treinamento e a calibração do modelo foram realizados seguindo a abordagem de validação cruzada do tipo k-fold (com k=10), utilizando 70% do conjunto de dados para essa finalidade (Kuhn; Johnson, 2013).

A fim de avaliar a confiabilidade dos resultados da modelagem na classificação, foram utilizadas amostras independentes do conjunto de treinamento, representando 30% do conjunto de dados total. Essas amostras foram avaliadas por meio de métricas de desempenho, como a curva ROC (Receiver Operating Characteristic) e o valor de AUC (Area Under the Curve) (Kuhn; Johnson, 2013).

A curva ROC e a métrica AUC são ferramentas comuns usadas na análise de desempenho de modelos de classificação binária. A primeira é um gráfico que mostra a taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos (especificidade) para a classificação (Kuhn; Johnson, 2013; Boehmke; Greenwell, 2019). A segunda, área sob a curva ROC (AUC), é uma métrica que fornece uma medida de quão bem o modelo é capaz de distinguir entre as classes positivas e negativas, com valores variando de 0 a 1, em que 0 indica um modelo completamente inútil e 1 indica um modelo perfeito. Quanto maior o valor de AUC, melhor é o desempenho do modelo na predição de classificação binária (Kuhn; Johnson, 2013; Boehmke; Greenwell, 2019). A linguagem R foi a escolhida para a prática deste trabalho e a interface utilizada foi a RStudio.

Para a análise da importância das variáveis na predição de locais adequados para a instalação de usinas de biogás derivadas da cana-de-açúcar, aplicou-se uma função que quantifica os pesos atribuídos a cada variável durante o desenvolvimento de cada algoritmo na modelagem (Kuhn; Johnson, 2013). Esses coeficientes são normalizados em uma escala de 0 a 100 para cada cenário utilizado, sendo que valores mais elevados indicam maior importância das variáveis no processo de modelagem (Kuhn; Johnson, 2013).

A modelagem dos diferentes algoritmos pode ser aplicada para o conjunto de dados raster para ter a predição de locais menos e mais favoráveis, em probabilidade de ocorrência, à instalação de usinas de biogás derivado da cana-de-açúcar, com valores de pixel variando entre 0 e 1, respectivamente. Quanto ao raster da modelagem, foi aplicada uma máscara sobre áreas urbanas e de vegetação nativa, considerando o dado de uso e cobertura da terra.

Para fins de simulação, o resultado do melhor modelo foi aplicado para o estado de Goiás. A escolha do estado de Goiás é devido à sua segunda posição no ranking de produtores de cana-de-açúcar, ficando atrás apenas do estado de São Paulo (IBGE, 2020). A base cartográfica foi a mesma, exceto os trechos de hidrografia, dado que foi considerada a base cartográfica contínua do estado de Goiás (IBGE, 2022). Para o estado de Goiás, os dados geográficos matriciais foram reamostrados para 90m de tamanho de pixel, devido a maior área em relação ao estado de São Paulo para redução do esforço computacional no momento da predição do melhor modelo. Esses dados também foram projetados em coordenadas métricas na projeção cônica equivalente de Albers.

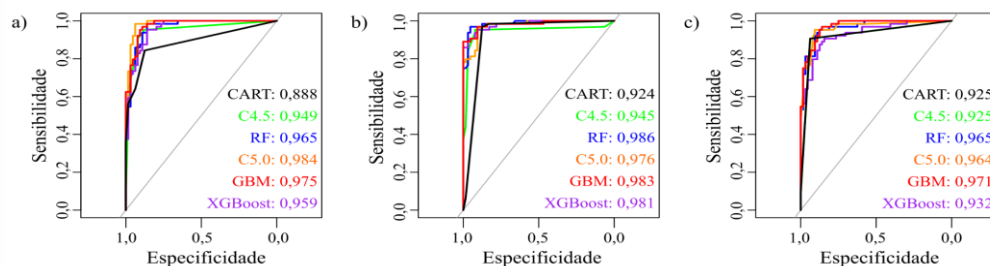
Complementarmente, foi realizada uma análise com base na localização das usinas termelétricas a biomassa, em operação que utilizam a cana-de-açúcar como fonte de biomassa em Goiás (EPE, 2023). Essa

escolha se justifica pelo estado ter apenas uma usina de biogás a partir de resíduos da cana, localizada no município de Goianésia, e pela sua relevância, uma vez que a biomassa da cana-de-açúcar é comumente utilizada tanto na produção de energia elétrica quanto na produção de biogás. Além disso, essa validação é especialmente viável, uma vez que as usinas termelétricas a biomassa focam na queima do bagaço de cana-de-açúcar, enquanto a produção de biogás envolve a utilização da torta de filtro e vinhaça, subprodutos que não são aproveitados por essas usinas termelétricas. Portanto, ao analisar as localizações das usinas de biomassa, podemos identificar áreas adequadas para a implementação de usinas de biogás, maximizando a utilização dos recursos disponíveis na agroindústria canavieira em Goiás.

Resultados

A análise do desempenho dos algoritmos em três cenários distintos, com base na métrica AUC, revelou resultados de notável relevância (Figura 2). Os cenários foram avaliados separadamente para identificar o desempenho superior de cada algoritmo em cada contexto específico.

Figura 2: Comparação dos desempenhos de algoritmos de aprendizado de máquina (CART, C4.5, Random Forest, C5.0, GBM, e XGBoost) com base na curva ROC para a) cenário 1; b) cenário 2; e c) cenário 3.



Fonte: Autoria própria.

No cenário 1 (Figura 2a), observou-se que o algoritmo C5.0 obteve a maior AUC, atingindo um valor de 0,984. O Random Forest também

apresentou um desempenho notável com uma AUC de 0,965. O algoritmo GBM e o XGBoost tiveram desempenho ligeiramente inferior, com AUCs de 0,975 e 0,959, respectivamente. O CART e o C4.5 tiveram as AUCs mais baixas neste cenário, com valores de 0,888 e 0,949, respectivamente.

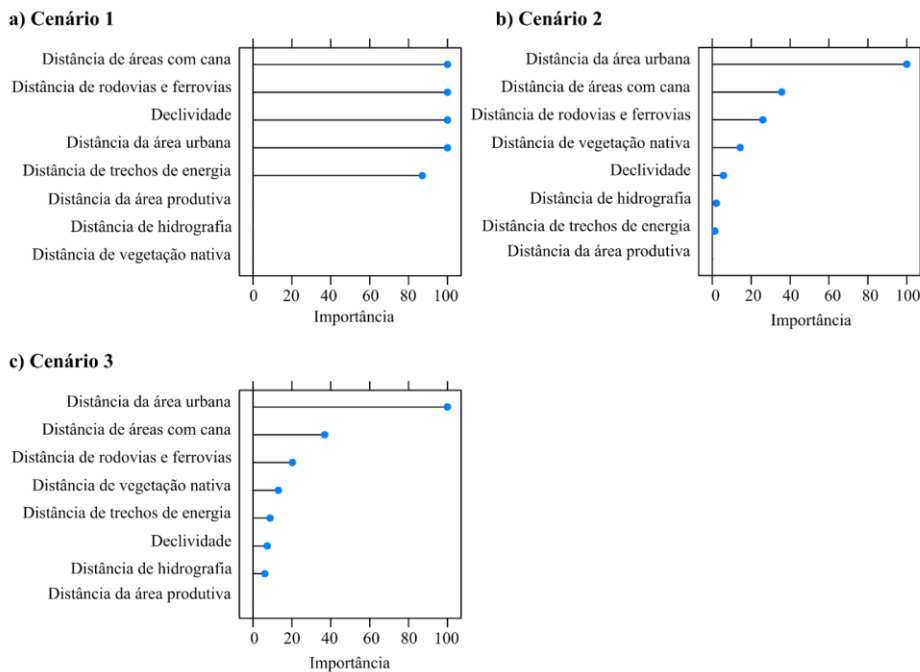
No cenário 2 (Figura 2b), o algoritmo Random Forest demonstrou o melhor desempenho, alcançando uma AUC de 0,986, a mais alta entre todos os algoritmos em todos os cenários. O C5.0 também obteve uma AUC elevada de 0,976, seguido pelo GBM com 0,983. O XGBoost e o C4.5 também apresentaram desempenho sólido com AUCs de 0,981 e 0,945, respectivamente. O CART obteve a AUC mais baixa neste cenário, com um valor de 0,924.

No cenário 3 (Figura 2c), o algoritmo GBM se destacou como o mais eficaz, com uma pontuação AUC de 0,971. O GBM demonstrou uma capacidade significativa de discriminação entre as classes positivas e negativas neste cenário, posicionando-o como o melhor algoritmo para esta situação específica. Tanto o algoritmo C5.0 quanto o Random Forest obtiveram AUCs de 0,964 e 0,965, respectivamente, demonstrando desempenho semelhante e competitivo. O XGBoost teve AUC ligeiramente mais baixa em comparação com os algoritmos mencionados, com valores de 0,932. O CART apresentou a AUC mais baixa novamente, com um valor de 0,925, juntamente com o C4.5.

Dessa forma, com base nos resultados AUC é possível concluir que o algoritmo C5.0 é o mais eficaz no cenário 1, o Random Forest é o mais eficaz no cenário 2, e o GBM é melhor em relação aos demais algoritmos analisados no cenário 3 (Figura 2). Essa análise ressalta a importância de adaptar a escolha do algoritmo ao contexto específico e aos objetivos da aplicação, bem como a necessidade de considerar o cenário em que a análise é realizada para garantir o melhor desempenho.

Em relação às posições de importância das variáveis (Figura 3) para os três cenários, têm-se destaque para a “Distância de Área Urbana”, “Distância de Áreas com Cana-de-Açúcar” e “Distância de Rodovias e Ferrovias”, com menos evidência para a “Distância de Área Produtiva”. A “Declividade” e a “Distância de Trechos de Energia” tiveram uma considerável importância para o cenário 1 em relação aos cenários 2 e 3, assim como o contrário ocorreu com a variável “Distância de Vegetação Nativa”, que obteve igual importância para os cenários 2 e 3 e nenhuma importância para o cenário 1. A “Distância de Hidrografia” não foi considerada com grande importância para os três cenários (Figura 3).

Figura 3: Importâncias das variáveis para a predição de usinas de biogás derivadas de resíduos de cana-de-açúcar nos três cenários. a) algoritmo C5.0; b) algoritmo Random Forest; e c) algoritmo GBM.

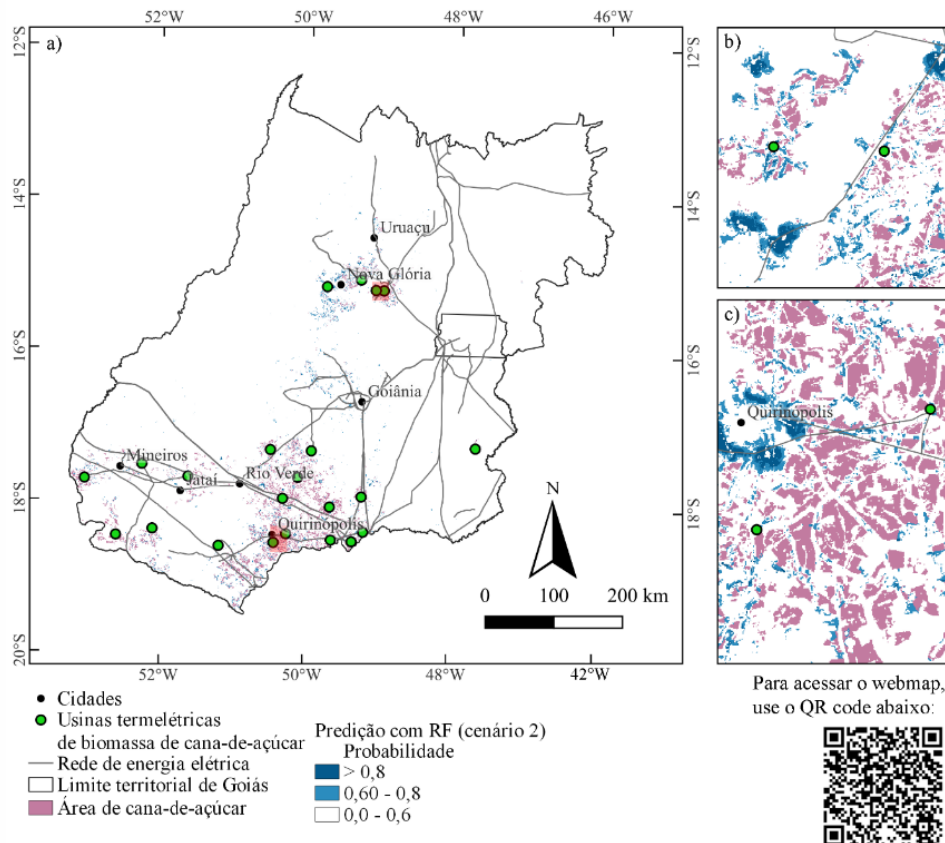


Fonte: Autoria própria.

Na Figura 4, apresentada no contexto do cenário 2, é exibida a predição dos locais propícios para a instalação de usinas de biogás derivadas de cana-de-açúcar no estado de Goiás, tendo como base o

algoritmo Random Forest. A legenda associada a essa predição atribui valores graduados que variam de 0 (representados em branco) a 1 (indicados em azul), sendo que as áreas em tons de rosa correspondem às regiões de cultivo de cana-de-açúcar. Observa-se que os locais mais propícios para a alocação de tais usinas estão situados em proximidade das áreas de cultivo de cana, das vias de transporte e das áreas urbanas. Este padrão sugere que a distância em relação à infraestrutura de transporte e à disponibilidade de resíduos de cana-de-açúcar emergem como as variáveis de maior relevância no processo de determinação da localização adequada para as usinas de biogás derivadas da cana-de-açúcar.

Figura 4: Produto cartográfico da predição para o estado de Goiás (a). Destaque para a região de Nova Glória e Goianésia (b), onde o ponto verde a esquerda corresponde a usina de biogás derivada de cana-de-açúcar em operação no território goiano, e na região de Quirinópolis (c).



Fonte: Autoria própria.

Na avaliação da aplicação do modelo preditivo de locais favoráveis à instalação de usinas de biogás derivadas de resíduos da cana-de-açúcar em Goiás, observou-se que, das 23 usinas termelétricas em Goiás que utilizam a cana-de-açúcar como fonte de biomassa, 20 apresentaram valores acima de 0.50 de probabilidade. Isso indica que o modelo preditivo treinado com dados referentes ao estado de São Paulo e aplicado a Goiás demonstrou replicabilidade em sua capacidade de identificar áreas propícias à instalação de usinas de biogás.

Considerando áreas com mais de 80% de probabilidade para o modelo de predição do cenário 2 ao utilizar o algoritmo Random Forest, foi identificado que 94 dos 246 municípios do estado de Goiás possuem locais favoráveis para a instalação de usinas de biogás derivado de cana-de-açúcar. As cidades melhor qualificadas estão concentradas na mesorregião do sul goiano, destacando municípios como Quirinópolis, Itumbiara, Cachoeira Dourada, Rio Verde, Santa Helena de Goiás e Jataí. Além disso, é possível destacar, na mesorregião do centro goiano, os municípios de Nova Glória, Goianésia, São Luiz do Norte, Itapaci e Ipiranga de Goiás.

Discussões

Em contextos de modelagem em aprendizado de máquina, é importante ressaltar que o desempenho de algoritmos pode ser substancialmente influenciado pela disposição estratégica dos pontos de pseudo-ausência (Chefaoui; Lobo, 2008). Neste contexto, a disposição dos pontos de pseudo-ausência exerceu uma influência sobre o desempenho de diferentes algoritmos, levando os algoritmos C5.0, Random Forest e GBM a demonstrarem resultados superiores em cenários distintos, nos quais foram aplicadas diferentes configurações espaciais para esses pontos.

Os algoritmos baseados em ensemble, como Random Forest, C5.0, GBM e XGBoost tiveram melhor desempenho em comparação com os

algoritmos individuais, como CART e C4.5. Esses algoritmos de ensemble são projetados para combinar as predições de vários modelos, o que geralmente resulta em um desempenho mais robusto e preciso (Mesri; Tahseen; Ogla, 2021).

Observando os resultados da métrica AUC nos cenários apresentados, pode-se notar que o Random Forest obteve as AUCs mais altas no geral, o GBM também obteve AUCs competitivas, e o C5.0 manteve um desempenho sólido. Esses algoritmos de ensemble são conhecidos por reduzir o sobreajuste (overfitting), lidar bem com dados ruidosos e fornecer predições mais estáveis, tornando-os geralmente uma escolha sólida em muitas aplicações de aprendizado de máquina (Mesri; Tahseen; Ogla, 2021).

Informações precisas sobre a distância de transporte fornecem conhecimento valioso durante a etapa de planejamento, permitindo o cálculo refinado das emissões de CO₂ e dos custos relacionados ao transporte e ao investimento (Höhn et al., 2014). Quando se trata da utilização do biogás na geração de eletricidade e calor, as instalações de biogás alcançam seu melhor desempenho em áreas de maior densidade populacional (Höhn et al., 2014).

As variáveis de maior importância para os modelos de predição foram "Distância da Área Urbana", "Distância das Áreas com Cana-de-Açúcar" e "Distância das Rodovias e Ferrovias". Estudos demonstram que os custos de produção e transporte desempenham um papel crítico na determinação do sucesso ou fracasso das usinas de bioenergia (Costa et al., 2020; Jayarathna et al., 2020; Latterini et al., 2020). Nesse sentido, há uma dependência em relação à localização geográfica da matéria-prima, já que tanto a produção quanto o transporte de biomassa representam uma parcela significativa dos custos envolvidos (Costa et al., 2020; Jayarathna et al., 2020; Latterini et al., 2020).

Goianésia, uma das cidades mais bem qualificadas para a instalação de usinas de biogás, inaugurou em 29 de setembro de 2023 uma planta de cogeração a partir da vinhaça proveniente da usina sucroalcooleira Jalles Machado. É relevante destacar que essa usina localizada em Goianésia apresentou probabilidade de 0,73 no modelo preditivo. Esse valor destaca a assertividade do modelo na identificação de locais propícios para a instalação de usinas de biogás e confirma a viabilidade da escolha estratégica de Goianésia como local de implantação.

Para uma transição eficiente de fontes de energia, há a necessidade de infraestrutura e políticas governamentais favoráveis para apoiar sua adoção. Isso inclui investimentos em redes elétricas inteligentes, incentivos fiscais e regulatórios, e políticas de tarifas que permitam a integração de fontes renováveis de energia na rede elétrica, por exemplo.

Para um futuro com sustentabilidade, a energia renovável é essencial como uma alternativa limpa e eco-friendly aos combustíveis fósseis, contribuindo para a redução da poluição, emissões de gases de efeito estufa e mitigação das mudanças climáticas. Uma das principais vantagens é que é menos poluente e resultante de recursos naturais, com um ciclo de renovação ao longo da vida humana sem esgotamento. Além disso, ela se torna cada vez mais acessível e competitiva, e como fonte descentralizada proporciona maior segurança energética e empregos locais.

Apesar de todas essas vantagens, ainda há, entretanto, desafios a serem superados na adoção de energias renováveis. Dois deles são a escolha do recurso a ser utilizado e a tomada de decisão para locais suscetíveis à produção de energia renovável. Diferentes recursos renováveis têm diferentes potenciais de produção de energia em diferentes locais (Bravo; Casals; Pascua, 2007). Além disso, a escolha do recurso e a seleção do local podem afetar não apenas a quantidade de energia

produzida, mas também os custos envolvidos, a viabilidade técnica e a aceitação social (Ioannou et al., 2018).

A energia eólica é, por exemplo, mais apropriada em locais com ventos fortes e constantes (Castro, 2004), enquanto a energia solar é mais adequada em regiões com grande exposição à luz solar direta, visto que é desta maneira que é gerado o efeito fotovoltaico (Zilles et al., 2012). A hidroeletricidade é mais indicada em áreas com rios de alta vazão e queda (Scartazzini; Livi, 1999), e a biomassa tem maior potencial em locais com grande quantidade de resíduos agrícolas e florestais (Ioannou et al., 2018), como é o caso da cana-de-açúcar das agroindústrias sucroalcooleiras.

Necessidades energéticas podem ser supridas por meio da utilização da biomassa, seja pela combustão direta ou da sua conversão em combustíveis gasosos, líquidos e/ou sólidos via processos termoquímicos ou bioquímicos. Além do mais, é necessário levar em consideração os impactos ambientais e sociais de cada fonte de energia renovável (Lozano-García et al., 2020).

A decisão sobre a escolha do recurso e a seleção do local também é influenciada por outros fatores, como disponibilidade de infraestrutura, acesso à rede elétrica, custos de instalação e operação, demanda por energia renovável, além de políticas e regulamentações governamentais (Bravo; Casals; Pascua, 2007; Ioannou et al., 2018). Dessa forma, é necessária uma avaliação cuidadosa e equilibrada de diversos fatores para se tomar a melhor decisão quanto à escolha da fonte de energia renovável e sua localização. Nesse sentido, pesquisas como esta podem contribuir de forma significativa. Os resultados aqui obtidos listaram fatores que impactam diretamente em variáveis ambientais, econômicas e sociais.

Uma melhoria da metodologia proposta seria agregar a GeolA para além do planejamento de localização, atingindo também o dimensionamento da planta (Yalcinkaya, 2020). Estudos de otimização de

rota e localização poderiam melhorar o refinamento do resultado do modelo (Laasasenaho et al., 2019). Algoritmos de aprendizado de máquina também poderiam ser usados para a previsão da produção do biogás (Li et al., 2022) e a otimização da operação da usina após instalada (Maghami; Mutambara, 2023).

Considerações Finais

Como considerações finais deste trabalho é possível estabelecer:

- Os resultados obtidos neste estudo permitiram identificar locais propícios para a instalação de usinas de biogás provindo da cana-de-açúcar da agroindústria goiana;
- Dentre os algoritmos treinados, o Random Forest foi o que obteve melhor resultado (AUC igual a 0,986), considerando o cenário 2. O modelo treinado para uma região, o estado de São Paulo, pôde ser aplicado para outra região, o estado de Goiás;
- A utilização da técnica de GeolA mostrou-se eficiente para esta predição, visto que locais onde tem a presença de usinas de biomassa derivada de cana-de-açúcar estão contidos na predição feita pelo algoritmo;
- Para a instalação de empreendimentos, especialistas poderiam ir a campo fazer uma análise e constatar se realmente os locais apontados pela aprendizagem de máquinas são de fato propícios;
- A partir deste estudo, tendo o modelo já implementado, poderiam ser geradas predições com a mesma abordagem para qualquer outra região do Brasil;
- A técnica empregada pode facilitar a análise e tomada de decisão para avaliação exploratória de locais favoráveis à produção de

biogás de biomassa da cana-de-açúcar, permitindo, assim, a identificação de regiões onde potencialmente um biodigestor poderia ser instalado para maximizar o seu uso e minimizar o impacto ambiental e os custos de instalação.

Referências

BOEHMKE, Brad; GREENWELL, Brandon M. **Hands-on machine learning with R**. 1. ed. Flórida: CRC press, 2019. 484 p.

BRAVO, Javier Domínguez; CASALS, Xavier García; PASCUA, Irene Pinedo. Gis approach to the definition of capacity and generation ceilings of renewable energy technologies. **Energy policy**, Elsevier, v. 35, n. 10, p. 4879-4892, 2007.

CASTRO, Rui M.G. Introdução à energia eólica. **Lisboa: Portugal: Universidade Técnica de Lisboa**, 82 p., 2004.

CHEFAOUI, Rosa María; LOBO, Jorge Miguel. Assessing the effects of pseudo-absences on predictive distribution model performance. **Ecological modelling**, 210, 478-486, 2008.

COSTA, Fabrício Rodrigues; RIBEIRO, Carlos Antonio Alvares Soares; MARCATTI, Gustavo Eduardo; LORENZON, Alexandre Simões; TEIXEIRA, Thaisa Ribeiro; DOMINGUES, Getulio Fonseca; CASTRO, Nero Lemos Martins de; SANTOS, Alexandre Rosa dos; SOARES, Vicente Paulo; MOTA, Pedro Henrique Santos; TELLES, Lucas Arthur de Almeida; CARVALHO, José Romário de. Gis applied to location of bioenergy plants in tropical agricultural areas. **Renewable energy**, Elsevier, v. 153, p. 911-918, 2020.

CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). **Valor Bruto da Produção - VBP**. 2023. Disponível em: www.cnabrazil.org.br. Acesso em: 25 fev. 2023.

DEPARTAMENTO DE ÁGUAS E ENERGIA ELÉTRICA (DAEE). **Base cartográfica digital**, escala 1:50.000 - Projeto GISAT. São Paulo: DAEE, 2008.

DAGNALL, Steve; HILL, Jon; PEGG, David. Resource mapping and analysis of farm livestock manures — assessing the opportunities for biomass-to-energy schemes. **Bioresource technology**, Elsevier, v. 71, n. 3, p. 225-234, 2000.

EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Sistema de informações geográficas do setor energético brasileiro**, 2023. Disponível em: <https://gisepeprd2.epe.gov.br/WebMapEPE/>. Acesso em: 05 nov. 2023.

FRANCO, Camilo; BOJESEN, Mikkel; HOUGAARD, Jens Leth; NIELSEN, Kurt. A fuzzy approach to a multiple criteria and geographical information system for decision support on suitable locations for biogas plants. **Applied energy**, Elsevier, v. 140, p. 304-315, 2015.

HÖHN, Jukka; LEHTONEN, Eeva; RASI, Saija; RINTALA, Jukka. A geographical information system (gis) based methodology for determination of potential biomasses and sites for biogas plants in southern finland. **Applied energy**, Elsevier, v. 113, p. 1-10, 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Base cartográfica contínua do Brasil na escala de 1: 100.000**. BC100_GODF versão 2022. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/bases-cartograficas-continuas.html>. Acesso em: 02 set. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Produção Agrícola Municipal. **Tabela 5457 - Área plantada ou destinada à colheita, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias e permanentes**. In: IBGE. Sidra: sistema IBGE de recuperação automática. 2020. Disponível em: <https://sidra.ibge.gov.br/tabela/5457>. Acesso em: 23 out. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Produto Interno Bruto - PIB**, 2023. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 25 fev. 2023.

IOANNOU, Konstantinos; TSANTOPOULOS, Georgios; ARABATZIS, Garyfallos; ANDREOPOULOU, Zacharoula; ZAFEIRIOU, Eleni. A spatial decision support system framework for the evaluation of biomass energy production locations: Case study in the regional unit of drama, Greece. **Sustainability**, v. 10, n. 2, 2018. ISSN 2071-1050.

JANOWICZ, Krzysztof; GAO, Song; MCKENZIE, Grant; HU, Yingjie; BHADURI, Budhendra. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. **International journal of geographical information science**, v. 34, n. 4, p. 625-636, 2020.

JAYARATHNA, Lasinidu; KENT, Geoff; O'HARA, Ian; HOBSON, Philip. A geographical information system based framework to identify optimal location and size of biomass energy plants using single or multiple biomass types. **Applied energy**, Elsevier, v. 275, p. 115398, 2020.

JENDE, Oliver; ROSENFELDT, Sebastian; COLTURATO, Luis Felipe de Dornfeld Braga; GOMES, Felipe Correia de Souza Pereira; PLATZER, Christoph; SERAVAL, Tathiana Almeida; HOFFMANN, Heike; CABRAL,

Carolina Bayer Gomes; BURKARD, Thilo; LINNENBERG, Carsten; NAU, Daniel; PEREIRA, Amaro; MARIANI, Leidiane. Brasil. Secretaria Nacional de Saneamento Ambiental. Probiogás. **Barreiras e propostas de soluções para o mercado de biogás no Brasil** / Probiogás; organizadores, Ministério das Cidades, Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ). Brasília, DF: Ministério das Cidades, 2016. 74 p.: il. - (Desenvolvimento do mercado de biogás; 4). ISBN 978-85-7958-058-1.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. 2013. ed. New York: Springer, 2013. 600 p.

LAASASENAHO, K; LENSU, Anssi; LAUHANEN, R; RINTALA, J. Gis-data related route optimization, hierarchical clustering, location optimization, and kernel density methods are useful for promoting distributed bioenergy plant planning in rural areas. **Sustainable energy technologies and assessments**, Elsevier, v. 32, p. 47-57, 2019.

LATTERINI, Francesco; STEFANONI, Walter; SUARDI, Alessandro; ALFANO, Vincenzo; BERGONZOLI, Simone; PALMIERI, Nadia; PARI, Luigi. A gis approach to locate a small size biomass plant powered by olive pruning and to estimate supply chain costs. **Energies**, MDPI, v. 13, n. 13, p. 3385, 2020.

LI, Chao; HE, Pinjing; PENG, Wei; LÜ, Fan; DU, Rui; ZHANG, Hua. Exploring available input variables for machine learning models to predict biogas production in industrial-scale biogas plants treating food waste. **Journal of cleaner production**, Elsevier, v. 380, p. 135074, 2022.

LOZANO-GARCÍA, Diego Fabián; SANTIBAÑEZ-AGUILAR, José Ezequiel; LOZANO, Francisco J; FLORES-TLACUAHUAC, Antonio. Gis-based modeling of residual biomass availability for energy and production in Mexico. **Renewable and sustainable energy reviews**, Elsevier, v. 120, p. 109610, 2020.

MAGHAMI, Mohammad Reza; MUTAMBARA, Arthur Guseni Oliver. Challenges associated with hybrid energy systems: An artificial intelligence solution. **Energy reports**, Elsevier, v. 9, p. 924-940, 2023.

MESRI, Khitam; TAHSEEN, Israa; OGLA, Raheem. Default on a credit prediction using decision tree and ensemble learning techniques. **Journal of physics: conference series**, 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). **Transforming our world: the 2030 agenda for sustainable development**, 2015. Disponível em: <https://sdgs.un.org/2030agenda>. Acesso em: 05 nov. 2022.

OPEN STREET MAP (OSM). **OpenStreetMap database [PostgreSQL via API]**. OpenStreetMap Foundation: Cambridge, UK; 2023.

ROMERO, Cristhy Willy da Silva; MIYAZAKI, Marcelle Rose; BERNI, Mauro Donizeti; FIGUEIREDO, Gleyce Kelly Dantas Araújo; LAMPARELLI, Rubens Augusto Camargo. A spatial approach for integrating gis and fuzzy logic in multicriteria problem solving to support the definition of ideal areas for biorefinery deployment. **Journal of cleaner production**, Elsevier, p. 135886, 2023.

SCARTAZZINI, Luiz Sílvio; LIVI, Flávio Pohlmann. Potencial hidroenergético do alto rio pelotas. **Revista brasileira de recursos hídricos**, v. 4, n. 4, p. 87-95, 1999.

SILVA, Sandra; ALÇADA-ALMEIDA, Luís; DIAS, Luís C. Biogas plants site selection integrating multicriteria decision aid methods and gis techniques: A case study in a portuguese region. **Biomass and bioenergy**, Elsevier, v. 71, p. 58-68, 2014.

SLIZ-SZKLINIARZ, Beata; VOGT, Joachim. A gis-based approach for evaluating the potential of biogas production from livestock manure and crops at a regional scale: A case study for the kujawsko-pomorskie voivodeship. **Renewable and sustainable energy reviews**, Elsevier, v. 16, n. 1, p. 752-763, 2012.

SOUZA, Carlos; SHIMBO, Julia; ROSA, Marcos; PARENTE, Leandro; ALENCAR, Ane; RUDORFF, Bernardo; HASENACK, Heinrich; MATSUMOTO, Marcelo; FERREIRA, Laerte; SOUZA-FILHO, Pedro; OLIVEIRA, Sergio; ROCHA, Washington; FONSECA, Antônio; MARQUES, Camila; DINIZ, Cesar; COSTA, Diego; MONTEIRO, Dyeden; ROSA, Eduardo; VÉLEZ-MARTIN, Eduardo; WEBER, Eliseu; LENTI, Felipe; PATERNOST, Fernando; PAREYN, Frans; SIQUEIRA, João; VIERA, José; NETO, Luiz Ferreira; SARAIVA, Marciano; SALES, Marcio; SALGADO, Moises; VASCONCELOS, Rodrigo; GALANO, Soltan; MESQUITA, Vinicius; AZEVEDO, Tasso. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote sensing**, v. 12, n. 17, 2020.

SULTANA, Arifa; KUMAR, Amit. Optimal siting and size of bioenergy facilities using geographic information system. **Applied energy**, Elsevier, v. 94, p. 192-201, 2012.

YALCINKAYA, Sedat. A spatial modeling approach for siting, sizing and economic assessment of centralized biogas plants in organic waste management. **Journal of cleaner production**, Elsevier, v. 255, p. 120040, 2020.

ZHAO, Bingchao; WANG, Han; HUANG, Zhihao; SUN, Qianqian. Location mapping for constructing biomass power plant using multi-criteria decision-making method. **Sustainable energy technologies and assessments**, Elsevier, v. 49, p. 101707, 2022.

ZILLES, Roberto; MACÊDO, Wilson Negrão; GALHARDO, Marcos André Barros;



Dados geográficos e inteligência artificial na predição de locais favoráveis para a instalação de usinas de biogás da agroindústria canavieira

Marlísia D'Abadia de Pina. Édipo Henrique Cremon

OLIVEIRA, Sérgio Henrique Ferreira de. **Sistemas fotovoltaicos conectados à rede elétrica**. 1. ed. São Paulo: Oficina de textos, 2012.

ZURELL, Damaris; FRANKLIN, Janet; KÖNIG, Christian; BOUCHET, Phil Jean-François; DORMANN, Carsten; ELITH, Jane; FANDOS, Guillermo; FENG, Xiao; GUILLERA-ARROITA, Gurutzeta; GUIBAN, Antoine; LAHOZ-MONFORT, José; LEITÃO, Pedro; PARK, Daniel; PETERSON, Townsend; RAPACCIUOLO, Giovanni; SCHMATZ, Dirk; SCHRÖDER, Boris; SERRA-DIAZ, Josep; THUILLER, Wilfried; YATES, Katherine; ZIMMERMANN, Niklaus; MEROW, Cory. A standard protocol for reporting species distribution models. **Ecography**, Wiley Online Library, v. 43, n. 9, p. 1261-1277, 2020.

Publisher

Universidade Federal de Goiás. Instituto de Estudos Socioambientais. Programa de Pós-graduação em Geografia. Publicação no Portal de Periódicos UFG.

As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

Contribuições dos autores

Todos os autores ofereceram substanciais contribuições científicas e intelectuais ao estudo. As tarefas de concepção e design do estudo, preparação e redação do manuscrito, bem como, revisão crítica foram desenvolvidas em grupo. O primeiro autor Marlísia D'Abadia de Pina ficou especialmente responsável pela concepção, metodologia, aquisição de dados e suas interpretações e análise, software, validação e curadoria dos dados. O segundo autor e autor correspondente Édipo Henrique Cremon pelo desenvolvimento teórico-conceitual, análise formal, investigação e validação. Todos os autores contribuíram para a escrita do documento. Declaramos ainda ciência das Diretrizes Gerais do BGG.

Marlísia D'Abadia de Pina, Bacharela em Engenharia Cartográfica e de Agrimensura e Mestra em Tecnologia, Gestão e Sustentabilidade pelo Instituto Federal de Educação, Ciência e Tecnologia de Goiás. Atua principalmente nos seguintes temas: Sensoriamento remoto, SIG, Machine learning e fontes de energia renovável.

Édipo Henrique Cremon, Geógrafo pela Universidade Estadual de Maringá. Mestre e Doutor em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais-INPE, com período de doutorado sanduíche na University of Exeter (Reino Unido). Atualmente é Professor Efetivo do Instituto Federal de Educação, Ciência e Tecnologia de Goiás-IFG (Campus Goiânia) e pesquisador do Grupo de Estudos em Geomática (GEO). Atua principalmente nos temas: Machine learning aplicado a dados geográficos para análise ambiental e geomorfológica.

Data de recebimento 20 de novembro de 2023

Aceite em 22 de janeiro de 2024

Publicação em 24 de maio de 2024